LAMOST数据流与光谱质量控制研究*

郭炎鑫[1,2]，罗阿理[1]

(1. 中国科学院光学天文重点实验室（国家天文台），北京 100101；2. 中国科学院大学，北京 100049)

摘 要： 郭守敬望远镜(LAMOST)每日产生海量观测数据，数据处理涉及观测计划生成、二维和一维数据分析、参数测量、质量控制和光谱释放等诸多环节。为了更高效地获取、处理、分析数据和发布数据以及及时解决数据处理过程中出现的问题,开展了LAMOST数据流与光谱质量控制研究。首先，深入研究LAMOST系统数据流和工作流程，结合关系型数据库进行数据建模，实现基于Linux的MySQL数据库系统，将数据处理和发布各个环节有机串联并融合在一起；然后，基于该数据库系统，定义光谱质量控制模型，建立光谱质量控制系统，严格控制光谱质量和光谱产出各个环节，从而为优质的光谱资源的释放提供保障。该数据流与光谱质量控制系统可以很好的满足LAMOST数据处理和数据管理的需要，是可以扩展至同类望远镜系统进行数据处理的一种有效方案。

关键词： 天文数据处理；　数据库；　数据流；　质量控制

中图分类号： TN216　　**文献标识码：** A　　**文章编号：** 1007-2276-(2004)4-0338-05

郭守敬望远镜（LAMOST，大天区面积多目标光纤光谱天文望远镜）[1]是一架新类型的大视场兼备大口径望远镜。2017年6月，LAMOST圆满完成了为期五年的第一期光谱巡天任务，共获取了约900万条光谱数据，其中高质量光谱数（信噪比大于10）超过700万及超过500万组恒星参数，远远超过了全世界光谱巡天项目获取的光谱数总和，为天文学家研究银河系及一般星系的形成与演化提供了重要的基础性数据。

LAMOST每夜上万天体的观测和数G字节的FITS[2]构成了一个海量数据集合。为了高效地利用LAMOST的观测能力，LAMOST已经建成了一套完整的自动化观测、数据处理和存储的软件系统[3]，其中主要包括巡天战略系统(SSS)、观测控制系统(OCS)、二维光谱数据处理(2D)、一维光谱数据处理系统(1D)。 如何有效地维护和管理巡天过程中产生的海量观测和光谱数据，并将这些数据和以上软件系统有机地融合起来是光谱数据处理系统要解决的首要问题之一。

借助于数据库系统来对数据处理数据流进行整合和管理是一种行之有效的解决方案。国内外越来越多的望远镜巡天项目带来越来越多的海量数据存储和处理的客观要求，对数据库技术要求也越来越高。虽然国外已经建立了成熟的数据服务网站，比如 ADS、SDSS 等，虽然它们都能够提供某个方面的服务，比如数据交叉服务，数据可视化服务，数据下载服务等，但是它们都是在数据处理和发布之后的服务，专门针对天文数据处理整个过程中数据流的数据库系统还没有先例。这需要我们在充分了解数据处理流程及其过程中所涉及到的数据流情况的基础上，探索建立适合LAMOST 数据流的数据库系统，同时能够融合光谱质量控制机制进来，协助我们的数据存储和处理、发布。

1　数据处理流程分析

LAMOST已经独立建立了一系列数据处理软件，除了巡天战略系统(SSS)、观测控制系统(OCS)、二维光谱数据处理(2D)、一维光谱数据处理系统(1D)之外，数据质量控制还需要加入观

* 基金项目：国家自然科学基金资助项目（11703053）；国家重点基础研究计划（973 计划）（2014CB845700）
收稿日期：2018-10-12；修定日期：2018-10-12
作者简介：郭炎鑫，女，博士．研究方向：天文技术与方法. Email: yxguo@bao.ac.cn

测日志检查环节和人工光谱检查环节等。以下是LAMOST数据处理过程中涉及到的关键环节：

（1）巡天战略系统（SSS）：负责自动化地制定观测计划，决定观测时间和安排观测流程，充分利用LAMOST的观测能力有效的组织巡天观测，提高观测效率，缩短观测周期。每个观测夜大概需要6-8个观测计划，每个观测计划大概有3600个源。

（2）每日观测日志入库：在每天观测完成后，结合人工复核，自动化入库各个观测天区实时的天气等状况，标记天区类型包括: 科学观测与否、月相、数据是否有效等，后期的质量控制会根据标记选择不同的数据发布策略，比如对于无效数据将无法发布给用户。

（3）2D数据处理软件（2D Pipeline）[3]：二维光谱处理的对象是光谱的CCD图像，目标是将二维图像的流量抽取成一维光谱，并扣除本底流量、杂散光、宇宙线和天光的干扰，再利用定标灯进行波长定标，利用流量标准星进行相对流量定标，最后将同一个目标的多次曝光和红蓝端合并成最终的光谱，在此过程中还要进行例如平场改正在内的其他的改正。

（4）1D光谱分析软件（1D Pipeline）[4]：它的处理对象是2D数据处理输出的一维光谱，目的是对光谱进行分类，并测量天体的红移，然后对恒星进行细分类。最后，Pipeline生成的四个主要分类分别为 STAR,GALAXY,QSO 和 UNKNOWN。

（5）CFI[3]：为了提高LASP收敛速度和结果的准确性，相关函数初始值（CFI）测量首先会生成一组大气参的初值，作为后续LASP进行精确大气参数测量的范围，然后由Ulyss[6]生成最终的大气参数测量值。

（6）LASP[3,6]：LAMOST stellar parameter（LASP）pipeline是完整的大气参数测量程序。它专门针对分类为AFGK型恒星的且满足一定信噪比条件的光谱，用Ulyss程序自动测量给出他们的有效温度、表面重力、金属丰度和视向速度。

（7）星系测量软件[3]：星系和类星体的识别和红移的测量是一项艰巨的任务，受望远镜极限星等的影响，很多星系和类星体隐藏在低质量光谱中，因此1D pipeline在处理星系和类星体的光谱上还有一定的缺陷，而且往往。所以，在初始的1D光谱分析运行完之后，还需要一个额外的独立的用于星系识别及其红移的测量的pipeline。

（8）人工光谱检查：LAMOST系统数据流中包括两个人工检查环节，2D检查和1D检查，前者是为了发现定标星有无异常，数据有无杂散光等大批量的数据问题，后者是为了保证数据分类和红移的准确性。只要这两部分检查出现问题，就会反馈给2D或者1D进行重新处理，相当于数据流出现小循环。

（9）星表整合：在数据检查完毕以及数据处理和参数测量等一系列工作完成之后，会根据多项判据整合出满足数据发布条件的几个星表，以备数据发布时与信息中心对接。星表整合环节和人工检查环节共同配合完成光谱质量控制过程。

（10）发布数据打包：在星表整合完成之后，对于即将释放的光谱数据，需要根据数据发布要求重写FITS头，重新画缩略图，以备数据上线真正发布。

（11）数据产品发布：按照数据发布要求的节点，我们会将整合好的星表以及打包好的数据一并推送给信息中心，并完成相应的数据统计。

2 数据流设计与光谱质量控制要求

LAMOST数据处理流程复杂，涉及环节众多，需要一个系统级的核心数据库来将这些环节及其数据进行衔接。该数据库系统的用户包括LAMOST 巡天与数据部全体工作人员，重点是巡天战略系统SSS，观测控制系统OCS，二维数据处理系统以及一维数据处理系统开发人员，通过访问该数据库可以实时获取所需数据、分析处理结果，从而改进程序处理方法。最终使得LAMOST数据库能够达到以下目标:

（1）对于天文学家选定的观测目标,该数据库系统存储了从SSS生成观测计划、OCS 获取观

测图像、2D Pipeline 处理CCD图像数据以及1D Pipeline输出光谱产品各阶段的主要信息，以帮助天文学家追踪他们的数据，更高效地查询各个处理过程的中间结果和获取有用的光谱产品输出。

（2）Pipeline面临软件版本更新，这时候需要对上一版本处理过的数据进行再次处理，该数据库系统可以帮助数据处理程序开发人员存储和管理关键数据、查询查询处理结果，进而通过分析和比较这些结果对程序和处理算法做出改进。

（3）具有良好的可扩展性和兼容性，能够适应时刻变化的人工检查和数据处理流程，能够对原系统最小改动情况下实现新环节的系统级融合和提供高效的检索服务。

（4）本数据库系统一维数据处理结果数据库表与数据存储和发布部门进行对接，协助实现一套高效的数据发布系统。

光谱质量控制的目标就是及时发现光谱质量问题，既包括由杂散光、天气、仪器等原因导致的成批问题光谱，又包括在2D Pipeline数据处理过程中发生的错误导致的光谱问题。然而，伴随着将要对国内外天文学家释放的光谱，还有1D Pipeline光谱分析的结果，分类红移等。对于低质量光谱，1D Pipeline不能保证百分之百的正确率，所以也需要进行质量控制。LAMOST系统数据流中包括两个人工检查环节,2D Pipeline结果检查和1D Pipeline结果检查，一方面我们需要结合核心数据库进行检查结果的写入和查询，另一方面也需结合网络技术将人工检查环节变得更加高效和方便。图1所示为光谱质量控制具体的流程，它主要是分为恒星和星系/类星体两条主线来完成。

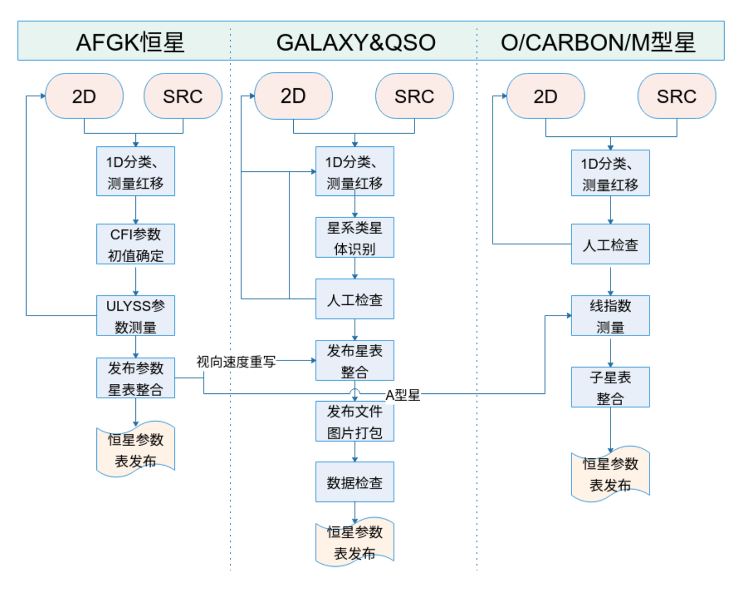


图1 LAMOST光谱质量控制流程图

Fig.1 Illustration of LAMOST spectral quality control

综上所述，我们的质量控制要求达到以下目标：

（1）2D Pipeline结果检查：快速浏览光谱，及时发现问题并录入数据库。

（2）1D Pipeline结果检查：对河外光谱，逐条检查分类和红移正确性，同时也会检查选源为河外而光谱分类结果是恒星的光谱部分；对于特殊类型的恒星光谱，逐条检查分类和红移正确性，结果也会入库。

（3）实时掌握数据处理各个环节处理进度：对原始数据来说就是原始数据推送情况，单次曝光缺失情况等；对pipeline来说就是处理进度，异常监控，系统负载等情况；对人工检查来说就是人工检查进度，任务分配等。

（4）光谱产品追溯：对单个目标，可以给出它自选源、观测、pipeline处理结果、信噪比、质量控制标记、生命周期和生命结束原因等的查询；对于一个天区，可以提供光谱整体质量统计分析，各环节输入输出检查；从时间角度讲，最终可以对特定时间点提供光谱所处状态的查询。

（5）数据流检查：各处理环节能够实时获取其输入输出结果，对数据库和文件系统及pipeline

进行自检和交叉检验，保证数据生产过程的可靠性。

（6）产品封装和上线：对光谱产品进行自动化的包括FITS头重写和打包在内的封装过程，同时实现与台网络中心进行数据库和数据对接，为每季度和每个正式对外释放数据提供一个便捷的数据库。

4　系统设计与实现

要一并完成以数据库为核心的数据流设计和光谱质量控制系统，我们采用了MySQL+PHP+Apache的架构方案。开源关系型数据库MySQL为我们提供数据流存储，Apache网页服务器端用PHP开发的光谱人工检查与结果反馈系统为我们日常的光谱检查工作平台。数据库和光谱检查平台一同完成光谱质量控制。以下是所用软件：

（1）PHP5：开源、跨平台、服务器端嵌入式动态网页开发脚本语言，具有数据库访问速度快、运行效率高、性能稳定等优点，完全支持SQL标准，可以兼容绝大多数数据库系统；

（2）Apache：目前应用最广泛的web服务器软件，它支持多种操作系统，功能强大且完全免费；

（3）MySQL：快速、多用户、多线程的SQL数据库服务器软件。它支持标准的SQL语句，支持多种平台，提供多种客户程序接口，适用于中型规模应用，完全开源。

4.1　数据流设计

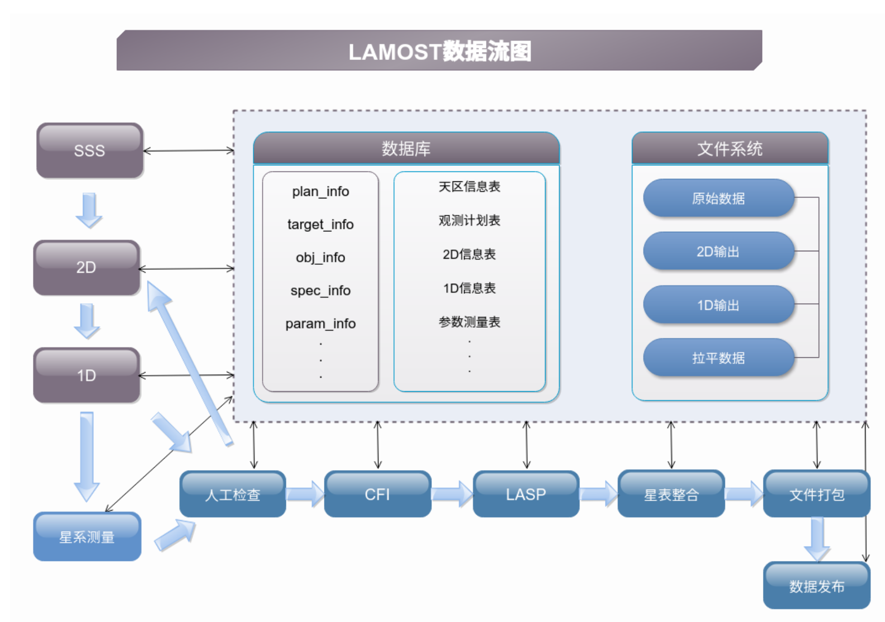


图2 LAMOST数据流图

Fig.2 Illustration of LAMOST dataflow

建立LAMOST数据处理系统离线数据库的目的是有效维护和管理数据处理过程中的数据流，在不同的数据处理阶段不同的软件承担着不同的职责：首先，SSS目标是为巡天观测制定观测计划，根据一个大的巡天星表，结合观测时的约束条件，寻找最佳位置，分配目标到光纤。这时需要将必要的统计信息，比如导星数量、目标数量、成功分配数量、分配天光数量、标准星数量以及标准星信息存入数据库以便2D程序访问。同时，2D处理过程中需要利用光纤分配的一些信息也需要在观测计划生成的同时写入该数据库。其次，OCS获取原始曝光图像之后不仅把数据打包以文件夹形式传回数据中心，而且要将每次曝光的信息，如曝光时间、曝光时长、光谱仪狭缝状态以及光谱仪温度等信息写入数据库中，这样2D在处理过程中才能获得充足的信息。然后，2D每处理一批原始数据，不仅将其处理的中间结果、最终结果存储在文件夹中，而且要将各个

目标的描述信息，包括目标对应的光谱仪号、光纤号、观测计划、所用pipeline版本号、处理日期、信噪比、输出文件夹路径等信息记录在数据库中。再次，1D通过查询数据库中2D输出的信息表记录来处理一维光谱数据，最终也将有价值的一维光谱信息写入数据库中。星系分类程序也会通过读取数据库中1D分类结果，将指定河外数据信息读出，待程序处理完毕将其结果写入数据库供人工检查阶段参考。最后，人工检查环节的检查结果也要保存下来，在后期的星表整合过程中使用。

图2为LAMOST数据流设计图。由数据库和文件系统将各个环节有机串联起来，每个处理环节都将会与文件系统或者数据库进行交互。通过查询数据库和文件可以实时掌握数据处理各个环节处理进度，追溯光谱产品，协助完成光谱质量控制。而对于人工检查环节来说，我们单独开发了一套软件，可以在线检查光谱，所有信息也会记录再后台数据库中。

4.2 数据库设计

在LAMOST尚未正式建成之前，[6]设计了基于MySQL/Linux的基本的数据库原型，[7]在先导巡天阶段根据实际数据处理情况完成了LAMOST的总体数据库设计和实现。在此基础上，我们优化设计了LAMOST正式的数据库系统。

该数据库系统由几个子数据库组成：数据日常处理数据库、成品数据库、光谱检查数据库。其中，数据日常处理数据库用于存放日常数据流中各个环节需要用到的信息；成品数据库用于存放整合好的可以对外释放的星表，而这里面又分为alpha版和正式版两个数据库；光谱检查数据库设计用来存放光谱人工检查相关信息和记录。整体的数据库结构示意图如图3所示。以下简要介绍各个数据库表的设计和结构：

*日常数据处理数据库：

(1) plan_info: 天区信息表，包括日期、天区名称、seeing、月相、是否科学天区标记等信息。一条记录对应一个plate。

(2) target_info: 选源信息表，存储选源信息表，一个记录对应一根光纤。一个plate的集合对应了SSS的相应观测计划。该表包括了目标位置、目标类型、选源来源、星等等信息。

(3) obj_info: 二维目标信息表，对应2D抽谱信息，包括mask编号、天区名称、光谱仪号、光纤号、曝光次数、数据版本、信噪比、抽谱完成存放路径、灯谱路径、平场路径等。

(4) spec_info: 一维光谱信息表，对应1D光谱分析结果。包括二维目标编号、1D软件版本号、光谱分类、红移、红移误差、置信度、1D光谱存放路径等信息。

(5) param_info: 待测参数的光谱信息表，包括满足测量大气参数的AFGK型恒星记录。提供1D光谱编号、1D光谱路径、参数测量标记、拉平光谱存放路径等信息。

(6) cfi_param: CFI信息表，包括了CFI测量的初始大气参数值即误差。

(7) cfi_param_norm: 光谱拉平后CFI信息表，包括了CFI测量的初始大气参数值即误差。

(8) uly_param: ulyss参数测量结果表，包括了ulyss精确大气参数测量结果值。

(9) uly_param_norm: 光谱拉平后ulyss参数测量结果表，包括了ulyss精确大气参数测量结果值。

(10) extragalaxy_info: 待测星系或类星体光谱信息表，包括了满足星系测量模块条件的所有光谱记录。提供1D光谱编号、1D光谱路径、信噪比、选源来源等信息。

(11) GQ_info: 星系测量结果信息表，包括了所有星系模块测量出来的星系或者类星体的类型、细分类、红移、置信度等信息。

(12) GM_info: 星系测量结果中的星系部分信息表，结构同GQ_info。

(13) QM_info: 星系测量结果中的类星体部分信息表，结构同GQ_info。

(14) fail2d: 2D处理失败光谱仪信息表，包括了失败光谱仪的日期、光谱仪号、天区名称、失败原因、软件版本等信息。

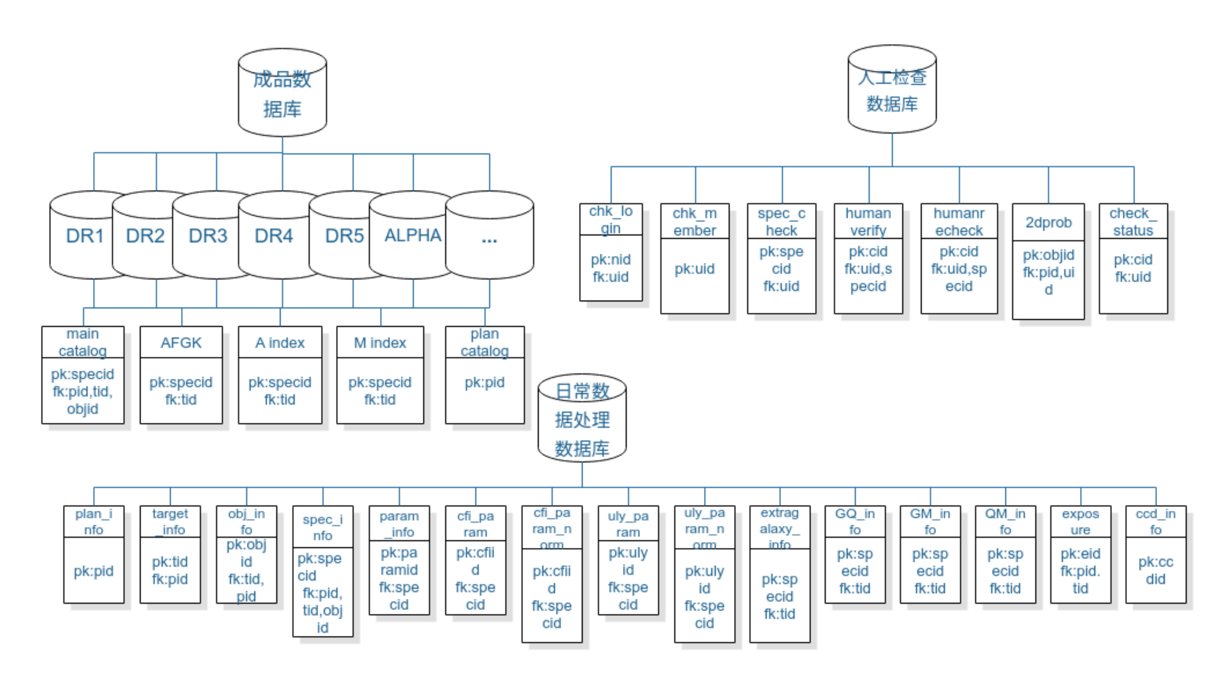


图3 数据库整体设计图

Fig.3 Designation diagram of the whole databases

(15) ccd_info: CCD状态信息表，包括了CCD的编号和增益。

(16) mask_info: MASK定义表，包括MASK编号，类型和含义。

(17) exposure: 曝光信息表，记录每次曝光的日期、开始时间、结束时间、曝光类型、狭缝状态等信息。

值得注意的是，这里仅仅是相同版本处理软件对应的表，如果从2D开始软件升级了，那么就会有一系列新的表产生，这样不同版本之间也可以相互比较和分析。

*成品数据库：

（1）DR*数据库：每个DR都有自己的数据库，目前从DR1到DR5共5个。

总光谱星表：整合好的所有光谱星表，包括了满足发布条件的部分，通过if_release字段是否为1可以卡出来。一条记录对应一条光谱，包括了选源信息、位置信息、分类、红移在内的所有关于这颗源的信息，在数据发布时，按需卡出子集和所需字段交付给信息中心。

-A型恒星星表：整合好的A型星参数星表，包括了每个DR所有A型星的线指数信息。

-M型恒星星表：整合好的M型星参数星表，包括了每个DR所有M型星的分子带指数信息等。

-AFGK型恒星高高质量量光谱参数星表：整合好的高质量的AFGK型恒星大气参数星表，包括了每个DR所有AFGK型恒星的有效温度，重力加速度，金属丰度，视向速度及其误差等信息。

-天区信息表：整合好的所有以观测天区信息表，一条记录对应一个plate，包括了中央星位置、seeing、曝光时间等信息。

（2）DR*alpha数据库：每年都有alpha版的数据于每个季度对外释放，所以每一年对应一个alpha数据库，截止到目前共有6个，即到DR6_ALPHA。而对于每个子数据库，他们所包含的数据表及结构都是一样的。

-总光谱星表：结构同DR*数据库中的总光谱星表一致。

-AFGK型恒星高高质量量光谱参数星表：结构同DR*数据库中的AFGK型恒星高高质量量光谱参数星表一致。

-天区信息表：结构同DR*数据库中的天区信息表一致。

这里没有A型恒星及M型恒星星表是因为数据发布策略规定alpha版数据只提供大气参数，不用测量线指数。

*光谱检查数据库：

（1）chk_login：记录加密后的用户名和密码信息。

（2）chk_member：记录用户信息，编号，部门，权限等。

（3）check_status：记录光谱检查状态，是否完成及完成量、百分比等。

（4）spec_check：记录需要人工检查的光谱的信息，比如fits文件读取路径等。

（5）humanverify：记录人工检查结果，包括分类、红移、是否有特殊问题等。

（6）humanrecheck：记录人工复核结果，内容主要是分类和红移。

（7）2dprob：记录用户反馈的原始数据问题，包括杂散光、负流量、天光等问题，这种情况下需要反馈给相关工作人员进行及时处理。

对于以上数据库的具体表字段设计，本文不一一赘述，在此仅展示光谱检查数据库的逻辑设计图，如下图所示，具体表信息见上文光谱检查数据库介绍。

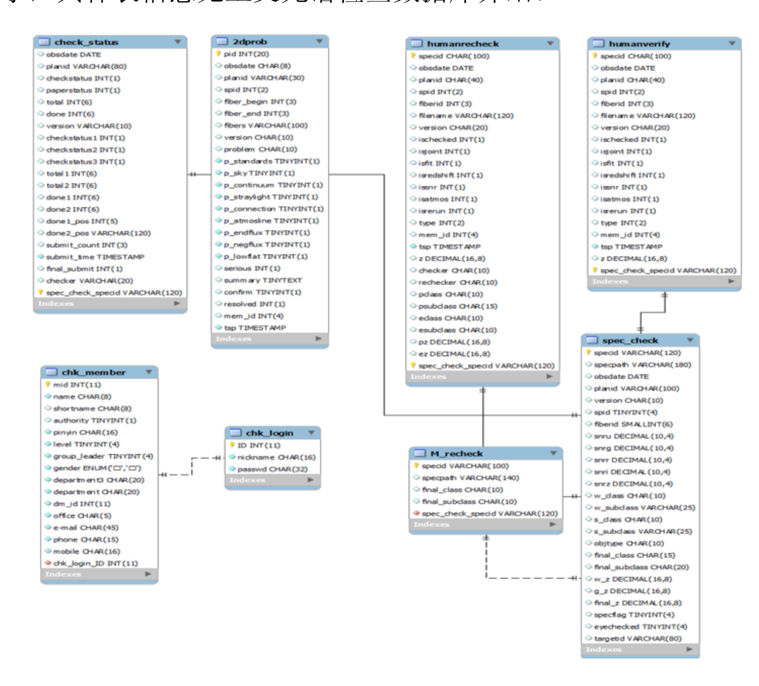


图4 光谱检查数据库表结构与关系图

Fig.4 Structure and Relationship Diagram of the spectra eye-check database

4.3 光谱检查与结果反馈系统实现

因为LAMOST可同时观测4000个天体，每个观测夜就能得到近3万条光谱，仪器效率的不均衡加上天气、视宁度的影响会导致数据质量即信噪比的降低，进而使得自动分类和测量参数的结果会不正确。与此同时，由于分类程序升级导致的同一批原始数据分类的结果不一致性也会存在。我们为了保证每一条光谱它的分类和测量信息的准确性，就需要人工经验或者说专家知识的介入，由于数据量大，逐条进行光谱人工检查是不切实际的，本软件就是为了最大程度的辅助专家进行自动化的光谱检查，通过该软件专家或者天文学家可以凭借自己的专家知识，对每条光谱标记出他认为正确的属性信息，对于由于观测原因导致的数据不可用情况也会予以反馈，所有信息将记录在后台数据库中。

4.3.1 所用软件及技术

该软件是基于BS模式的在线网页版的检查系统。服务器端需要用到基本的软件包括：

（1）MySQL存储后台数据库数据。

（2）PHP用于实现动态交互式网页。

（3）Python及相应的天文软件包和科学运算包，用于远程读取光谱数据的FITS文件。

（4）Apache网页服务器。

（5）一台内存32G、硬盘2T以上的高性能服务器，用于部署上述网页服务。

客户端只需要安装有浏览器就可以登录查看。

4.3.2 功能设计与实现

该系统设计为三个光谱检查部分，如图5所示，一个是星系或者类星体等比较容易错分的少量的光谱检查的页面（图5左上），一个是恒星等比较不易出错的光谱检查的页面（图5中上），同时还提供一个按照光谱仪进行2D预览检查的页面（图5右上）。该软件主要实现以下基本功能：

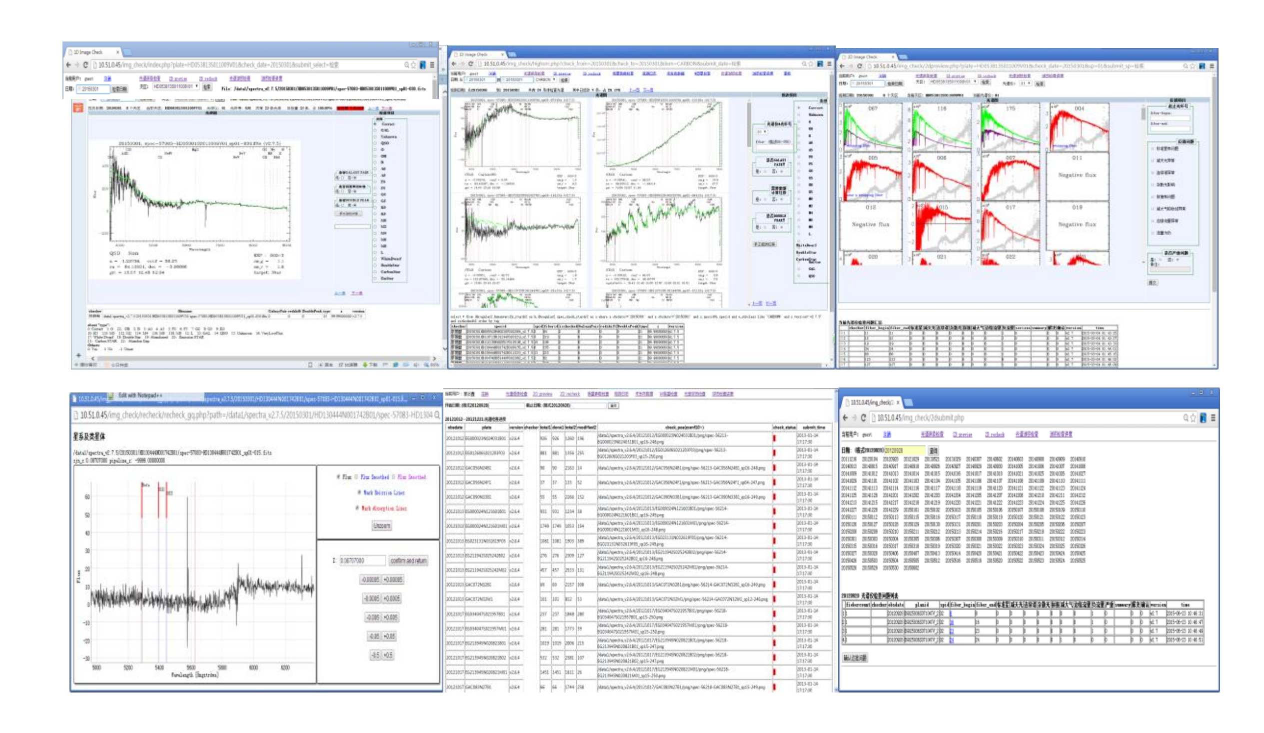


图5 光谱检查页面展示

Fig.5 Illustration of LAMOST spectra eye-check

（1）用户登录，密码验证。

（2）光谱逐条查看和结果提交，这是针对星系和类星体等河外光谱数据，需要最大化展示光谱图像供专家查看，给出正确的分类和红移结果，页面上提供细分类单选按钮供用户选择。

（3）光谱浏览查看和结果提交，这是针对恒星光谱数据，只需要成批的浏览查看，对其中的有误分类或者红移结果进行纠正。所有图片以日期-天区-光谱仪号-光纤号升序排序，点击其中任一幅图像，可以自动获取该图对应的id，点击提交之后可以记录目前查看的进度。

（4）按光谱仪成批检查原始数据或者2D处理是否有问题，并将其标记，如果是原始数据问题我们将丢弃它，如果是2D处理问题，我们会将标记自动反馈给2D相关人员进行重新处理。依次点击开始光纤号和结束光纤号文本框，然后点击图片，网页会自动识别光纤起止编号。当用户录入问题描述并点击提交按钮之后，该起止编号的问题光纤就会记录在数据库中。

（5）时间节点提交功能，浏览查看页面中记录用户上次查看的最后位置，以便于下次接着检查而无需从头查看，而且可以统计已经检查的数量，更好的掌握检查进度。

（6）红移测量，根据谱线匹配测量（图5左下）。该网页用Javascript控制光谱红移的增大和减小，实时展示光谱谱线的相应变化。

（7）检查进度查询，可以查询某些时间段的数据检查进度（图5中下）。

（8）2D Pipeline 检查结果响应，2D相关处理人员查询某些日期的用户曾经反馈的问题，确

认问题并在数据库中标记，直到查询结果空白就表示没有任何问题（图5右下）。

4.3.3 系统测试

将以上功能逐一单独测试，该软件系统都能顺利完成快速SQL读写，以及光谱的逐条检查或是浏览检查、红移测量等功能。

将该系统做负载能力测试，500个用户同时访问，系统均比较稳定。对于数据处理机房的32台刀片机，以及20多个用户来说这种稳定性和负载能力足够了。相信以后如果对服务器和硬盘等各种硬件环境升级的情况下，性能会有更大提升。

5 总结与展望

本文针对LAMOST数据处理特点，结合了数据库技术和网络技术，设计了LAMOST数据处理核心数据库和数据流，开发了一套光谱质量控制系统，第一次在兼顾光谱质量控制标准的同时，为大型望远镜后期数据处理阶段提供了数据库和数据流设计了基础模型。LAMOST二期巡天已经开始，中分辨数据会越来越多，在后续工作中，将进一步研究中分辨数据处理和发布需求，优化现有数据流和数据库，同时将质量控制扩展到中分辨数据处理中去。

参考文献：

［1］ Cui Xiangqun, Zhao Yongheng, Chu Yaoquan, et al. The Large Sky Area Multi-Object Fiber Spectroscopic Telescope (LAMOST) [J]. Research in Astronomy and Astrophysics, 2012, 12(9): 1197-1242.

[2] 崔辰州 , 李文，于策，徐祯 , 赵永恒 FITS数据文件的检索和访问[J]. 天文研究与技术, 2008 , 5 (2) :116-123

［3］Luo Ali, Zhao Yongheng, Zhao Gang, et al. The first data release(DR1) of the LAMOST regular survey [J]. Research in Astronomy and Astrophysics, 2015,15(8): 1095-1124.

［4］ 张昊彤，褚耀泉，陈建生. 多目标光纤光谱处理方法的研究[J]. 天文学报, 2001, 42(3): 313-328.

Zhang Haotong, Chu Yaoquan, Chen Jiansheng. Data reduction of the multi-object fiber spectra [J]. Acta Astronomica Sinica, 2001, 42(3): 313-328

［5］ 王凤飞. LAMOST一维光谱数据处理与分析软件的设计和实现[D]. 中国科学院国家天文台, 2009.

Wang Fengfei. Design and Implementation on spectra reduction and analysis software for LAMOST [D]. National Astronomical Observatories, Chinese Academy of Sciences, 2009

［6］ Wu Yue, Singh H P, Prugniel P, et al. Coudé-feed stellar spectral library – atmospheric parameters [J]. Astronomy & Astrophysics, 2010, 525(19): 297-304.

［7］ 李会贤，桑健，王沙，等. 基于mysql/LINUX的LAMOST数据库设计与实现 [J]. 天文研究与技术, 2006, 3(1): 56-63.

Li Huixian, Sang Jian, Wang Sha, et al. The database design of LAMOST based on MYSQL/LINUX [J]. Astronomical Research & Technology, 2006, 3(1): 56-63.

［8］ Guo Yanxin, Luo Ali, Wang Fengfei, et al. Design of LAMOST Data Processing and Production Database [C] //The International Society for Optical Engineering. Proceedings of SPIE, 2012, 8451.

Research of LAMOST Dataflow Designation and Spectral Quality Control

Guo Yanxin[1,2], Luo Ali[1]

(1. Key Laboratory of Optical Astronomy, National Astronomical Observatories, Chinese Academy of Sciences, Beijing 100101, China; 2. University of Chinese Academy of Sciences, Beijing 100049, China Email: lal@bao.ac.cn)

Abstract: Guo Shoujing telescope(LAMOST) produces massive observational data every day. Data processing involves a series of processes, such as observation program generation, two-dimensional data reduction and one-dimensional data analysis, parameter measurement, quality control and data releases. In order to obtain, process, analyze and release data more efficiently and solve problems happened during data processing in a timely manner, the research of the LAMOST data flow designation and spectral quality control were carried out. Firstly, we intensively study the LAMOST dataflow and workflow. By combining the relational database, data modeling is fulfilled and the database system is realized based on MySQL, and the data processing and releasing segments are integrated organically. Then, ground on the former system, we define the spectral quality control model and establish the spectral quality control system to strictly control the spectral quality and spectra output, so as to provide guarantee for the releases of high-quality spectral resources. This system can well meet the needs of LAMOST data processing and data management, and is an effective scheme that can be extended to similar telescope systems for data storage and processing.

Key words: Data Processing; Databases; Dataflow; Quality Control